

Final practical work: Run models, compare different model calibrations and compare your model outputs !

The aim of this exercise is to go through the different points that were evoked during the week from a practical point of view. Each of you will work independently on his specific subject and after obtaining the desired results, will be grouped (2 to 3 people) to compare your results.

A little presentation from each group of the model results (and an interpretation/discussion of these ones) will be asked and presented on Friday.

Case study: *Sterechnus neumayeri* (Meissner 1900)

S. neumayeri is a sea urchin with a circumpolar distribution. It has been proved to be an unique species with few ambiguities to be mismached with another specimen. It has an opportunistic feeding diet, being scavenger, detritivorous or predator. The sea urchin has been importantly studied during the past decades.

It has been chosen for this exercise because of the lack of genetic ambiguities and the abundance of occurrence data available (quite a rare case study in Southern Ocean ! ^).

Let's go with practice !

First of all, let's realise some collective changes and checkings in the run_yOur_SDM.R file :

1. Occurrence dataset

Use the provided file occurrences_sterechnus.csv

2. Change environmental dataset

Modify the dataset of the run_yOur_SDM.R document to integrate the following environmental descriptors.

1	Depth	9	POC_2005_2012_max
2	Geomorphology	10	Roughness
3	ice_cover_min	11	Sediments
4	ice_cover_max	12	seafloor_current_speed
5	ice_thickness_min	13	seafloor_sali_2005_2012_min
6	ice_thickness_max	14	seafloor_sali_2005_2012_max
7	mixed_layer_depth	15	seafloor_temp_2005_2012_min
8	POC_2005_2012_min	16	seafloor_temp_2005_2012_max
		17	Slope

These environmental descriptors were chosen for their ecological relevance and their availability for Southern Ocean benthic habitats.

Check the collinearity of the selected environmental descriptors and remove the layers that are strongly correlated

_ use Spearman correlation evaluation ('cor' function in R base library)

_ use the VIF approach ('usdm' library, threshold chosen at VIF :10)

How many layers should be removed ?

3. Realise little changes in the code

- Change the number of background data to 1000.

WHY ? ... In Barbet-Massin et al. (2012), they explain that for BRT algorithm, predictive deviance is reduced when the number of background data is close to the number of available presence data. However, the aim of the background data is also to represent the overall environmental conditions on which the model is projected. The number of background records should therefore be sufficient. Therefore, a trade-off number should be chosen.

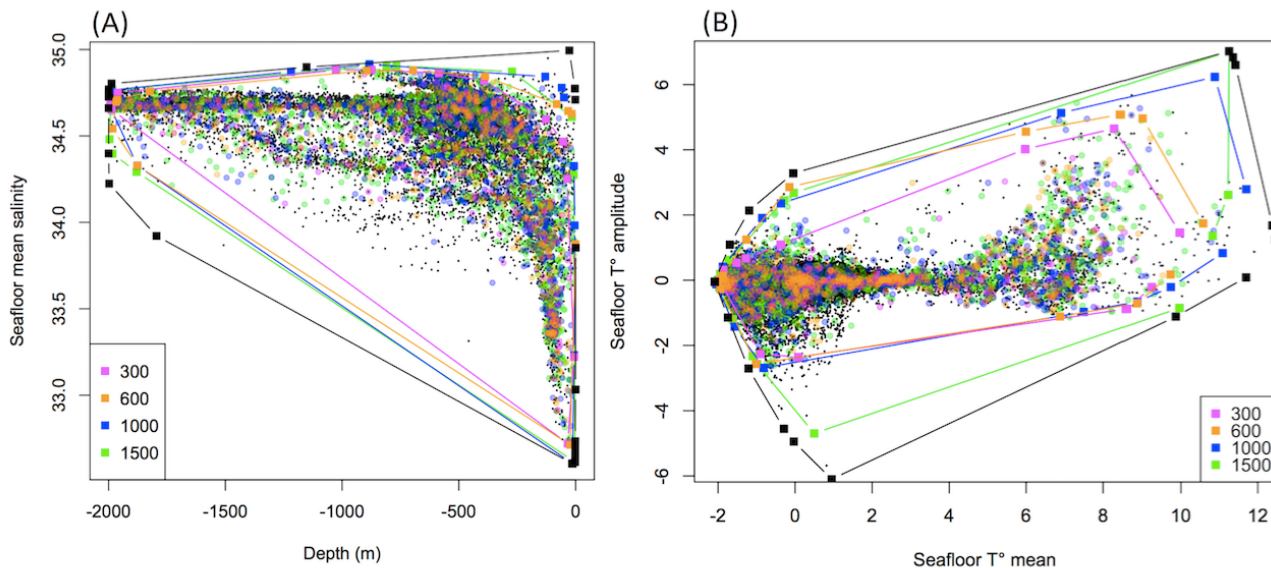


Figure : Evaluation of the number of random background data necessary to represent the all geographic environment. Representation of environmental values of the entire projected area at depth < 1500m (black dots) and environmental values with random sampling of 300, 600, 1000 and 1500 background data (coloured dots). Convex hulls were calculated using the 'chull' function of the *grDevices* R package; they delimit the environmental range covered by the background data samples. The number of points as close as possible of *Sterechinus neumayeri* presence records number while exhaustively representing the environment is preferred. 1000 data points was estimated to be the best option.

- Change the number of model replicates to 15. So that, each time you will generate a run, you will have results of 60 models. Have you understood why ?

Be careful, it will take around 5 minutes of computation to get a final model run and probably makes your computer go slower... don't forget to save your files regularly.

- Change the set of parameters to $tc=4$, $lr=0.01$, $bf=0.8$, which corresponds to the set of parameters that minimizes the predictive deviance. See Elith et al. (2008) for explanation about these parameters

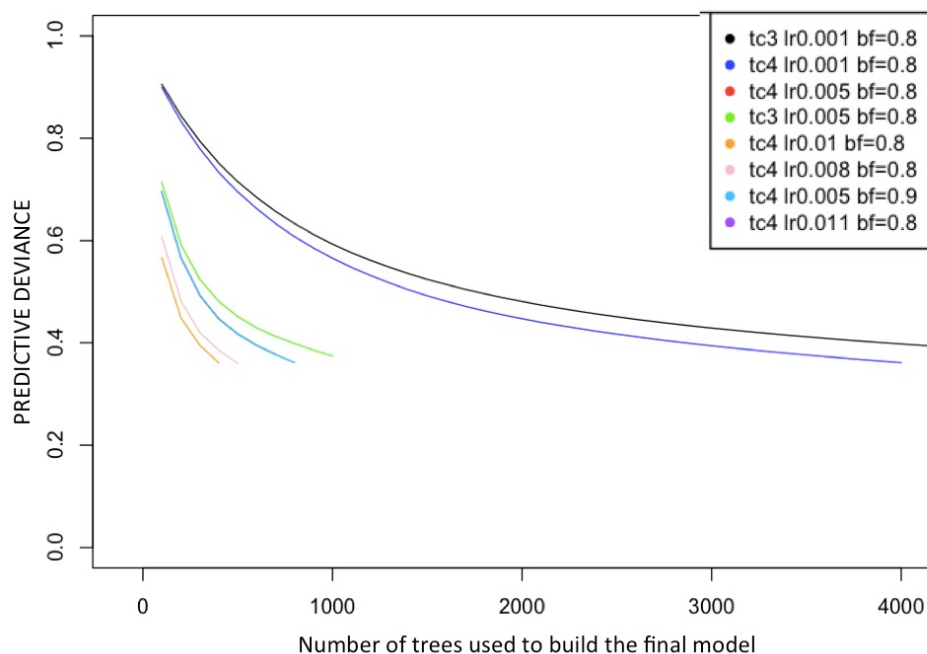


Figure : Trials of model and comparisons with different set of parameters. Figure computed with the *calibration_BRT.R* script

References (you can find them in the gitHub folder)

Elith, J., Leathwick, J. R., & Hastie, T. (2008). A working guide to boosted regression trees. *Journal of Animal Ecology*, 77(4), 802-813.

Barbet - Massin, M., Jiguet, F., Albert, C. H., & Thuiller, W. (2012). Selecting pseudo - absences for species distribution models: how, where and how many?. *Methods in ecology and evolution*, 3(2), 327-338.